ATTACHMENT 2

ATTACHMENT 2: Probabilistic Risk Assessments and Monte-Carlo Methods: A Brief Introduction

Risk assessments are a crucial part of EPA's pesticide regulatory program and have been for over 25 years. These assessments are used to estimate impacts on human health and the environment from the use of a given pesticide. Agency policy is that risk assessment should be conducted in a tiered approach, proceeding from simple to more complex analyses as the risk management situation requires. The Agency has traditionally used "deterministic" assessments involving point estimates of specific parameters to generate a single estimate of exposure and risk based on various assumptions about the concentration of pesticide in any given medium (e.g., food, water, air etc) and the amount of that medium consumed, breathed, or otherwise contacted. Deterministic assessments can begin with worst-case assumptions (for example, residues on foods at tolerance levels), then can be refined by more realistic values that remain point estimates (for example, average residues from field trials). Even with a tiered approach, each deterministic assessment provides single values for estimates of exposure from a given pathway. Such single-value risk estimates do not provide information on the variability and uncertainty that may be associated with an estimate.

Current Agency Policy (5/15/97) is that *probabilistic* analysis techniques (of which Monte-Carlo is one example) can be viable statistical tools for analyzing variability and uncertainty in risk assessments, provided they are supported by adequate data and credible assumptions. Probabilistic techniques can enhance risk estimates by more fully incorporating available information concerning the *range* of possible values that an input variable could take, and weight these values by their *probability* of occurrence. As an example, a particular food commodity (e.g., tomatoes) might contain a range of pesticide residues for any given pesticide, with a large percentage of tomatoes consumed actually containing no residues at all (since not all tomatoes are treated). In addition, individuals may or may not consume tomatoes on any given day and, over time, are expected to consume varying amounts of this food item due to varying daily consumption patterns. Probabilistic risk analysis permits OPP to assess the range of exposures (and their associated probabilities) which result from combinations of the various residue levels and consumption patterns. The resulting output of a probabilistic determination is a distribution of risk values with probability assigned to each estimated risk. Some of the major differences between deterministic and probabilistic estimates are summarized in the table below:

Deterministic Risk Assessment	Probabilistic Risk Assessment

• Pesticide concentrations and potential exposure factors are expressed as point estimates. • The risk estimate is also expressed as point value. The variability and uncertainty of this value are not reflected.	• Takes into account all available information and considers the *probability* of an occurrence. • The risk estimate is expressed as a distribution of values, with a probability assigned to each value. • The distribution reflects variability and uncertainty.

Tiered Approach to Risk Assessment

As risk assessments are refined, assumptions can proceed from more conservative (more health protective) to more realistic reflections of exposure. As noted above with the example of residues on food, such refinements can be applied to deterministic assessments. Probabilistic analyses, including Monte Carlo, represent numerical techniques to reflect more realistic assumptions. For example, Tier I of acute dietary assessments as conducted by OPP includes conservative assumptions such as: all foods consumed by an individual in any given day were treated with the pesticide in question (if registered for use on that food) and that residues are present in those consumed foods at the maximum legal limit. Monte-Carlo techniques fully applied to this situation would allow incorporation of information concerning the percent of the crop which is treated, the amount of pesticide applied and timing of its application, and the range and distribution of residue values expected to be found. This information is useful because a particular food (e.g., tomatoes) might contain a range of pesticide residues for any given pesticide, with a large percentage of tomatoes consumed actually containing no residues at all (since not all tomatoes are treated). In addition, individuals may or may not consume tomatoes on any given day and, over time, are expected to consume varying amounts of this food item due to varying daily consumption patterns. Any variability and uncertainty is explicitly included in the analysis and is fully disclosed.

The Origin of Monte-Carlo Techniques

Monte-Carlo techniques have been used since the 1940's when they were first developed by physicists working on the Manhattan project. Only recently, however, have personal computers become sufficiently powerful and widespread for Monte-Carlo techniques to be widely applied for health risk assessments. Modern spreadsheet programs now provide a range of critical facilities to help to illustrate and order a model including advanced statistical functions, charting, etc. And the simplicity and capabilities of recently introduced commercial Monte-Carlo software allows these techniques to become virtually all but routine.

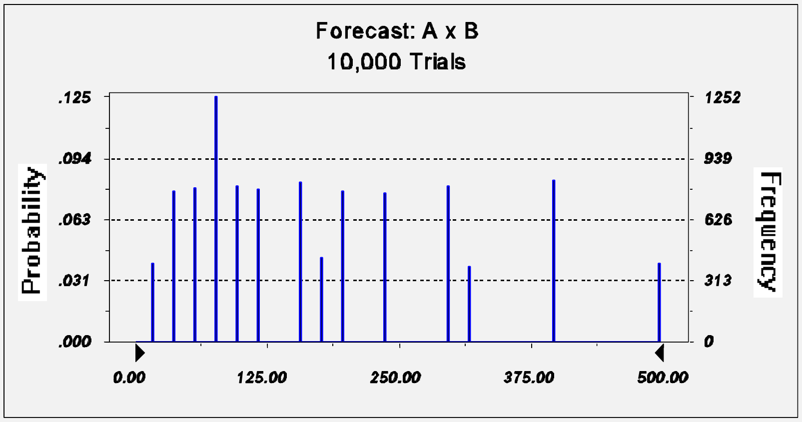
The origin of the name “Monte-Carlo” relates to the famous gambling city in Monaco, but the relation to gambling applies only to the probability of a given event occurring <u>over the long term.</u> Although one cannot know precisely which number will appear on the next roll of a craps die or the spin of a roulette wheel, one can predict over the long term (and as precisely as desired) the frequencies associated with each outcome. Monte-Carlo numerical techniques similarly cannot predict exactly which exposures will occur on any given day to any specific individual, but can predict the range of potential exposures in a large population and each exposure's associated probability.

What is Monte-Carlo Analysis?

Monte-Carlo analysis is simply one of several mathematical techniques for performing probabilistic risk assessments. The Monte Carlo technique, as applied to exposure assessment, involves combining the results of hundreds or thousands of random samplings of values from input probability distributions in such a manner as to produce an output distribution which reflects the expected range and frequency of exposures. Although computationally-intensive, Monte-Carlo techniques themselves are not complicated. Assessing a Monte-Carlo analysis requires examining the appropriateness of assumptions, judgements, and data sets which are key inputs to the mathematical procedures.

The first step in a Monte-Carlo simulation is the construction of a model that accurately represents the problem at hand. The makeup of the model usually entails a mathematical combination (addition, multiplication, logarithms, etc.) of the model input variables which can be expressed as probability distributions. If, for example, one desires to simulate the daily dietary pesticide exposure to individuals from a particular pesticide in tomatoes, this can be simulated by repeatedly drawing random values from two separate distributions: one distribution represents tomato consumption by individuals while the other represents pesticide levels in tomatoes. Here, the output variable (daily pesticide exposure) is defined as the product of the two input variables (tomato consumption in grams/day and pesticide residue concentrations in ug/g). Each random pair of input variables obtained from repeated independent samplings of the input distribution are multiplied together and the product used as one point in the distribution for the output variable. In general, this process is repeated thousands of times and the thousands of output products generated, taken together, form a distribution of frequencies. This technique is more fully illustrated in the box on the following page:

Suppose that our two input variables are defined as a and B where a = {2 , 4, 6, 8, 10}, B = {10, 20, 30, 40, 50}, and our output variable C is defined as the product of a and B (i.e., C = a x B). Set a might represent the concentration of a pesticide in tomatoes (in ug/g) and Set B might represent the daily consumption of tomatoes (in g/day). We wish to determine the range and frequency of potential values of C (which in this case would represent daily exposure to the pesticide in ug/day) . Inspection of the input data immediately reveals that the value for C (daily exposure) can range from a low of 20 ug/day (i.e., 2 x 10) to a high of 500 ug/day (i.e., 10 x 50) and that each of these two extreme values should occur approximately 4% of the time (i.e., 1/5 x 1/5 = 1/25 = 4%), Monte-Carlo methods permit us to evaluate **all** values that can be generated for the value C along with each of their associated probabilities. The Monte-Carlo method randomly chooses a single pesticide concentration value from Set a and a single tomato consumption value from Set B. These two values are multiplied together (to give daily pesticide exposure, C) and this resultant value stored. This process is repeated thousands of times with all values of C eventually plotted as a frequency histogram as shown above. Note that the lowest value is 20 ug/day and the highest value is 500 ug/day, just as originally predicted. Note also that these two values each occur approximately 4% of the time, just as (again) predicted from our original inspection. Although this example uses discrete values for sets a and B, Monte-Carlo modeling can also be performed when the input variable are described as continuous variables.



Regardless of how accurately the fitted distribution conforms to the data, or what method of sampling is used, the analyst has to set up a model that reflects the situation being assessed. According to Vose's *Quantitative Risk Analysis: A Guide to Monte Carlo Simulation Modeling*, the cardinal rule of risk analysis modeling is: "Every iteration of a risk analysis model must represent a scenario that could physically occur." Following this rule will lead to a model that is both accurate and realistic. As an example, it would be improper to model a cow diet as a random sampling of feeds with established tolerance for the pesticide of interest since many of the diets generated in such a manner would be unreasonable with respect to the roughage/nonroughage components, carbohydrate/protein mix, commodity combinations, and economic constraints. In short, blind application of Monte-Carlo techniques without regard for the reality of the generated scenarios will produce absurd results with no basis in reality. The analyst should ensure that each of the hundreds or thousands of iterations is a scenario with real-world possibilities.

It is often tempting in risk analysis modeling to include very unlikely events that would have a very large impact should they occur. A rare event of concern is defined as an event that has a low probability of occurrence but a potentially high impact on the results of a risk analysis. The expected impact of a rare event is determined by two factors: the probability that it will occur and the distribution of possible impacts. For example, widespread systematic illegal use of a pesticide or gross calibration errors in a pesticide's application might be a situation which occurs to some unknown (but relatively insignificant) extent. Since the probabilities of these events are so difficult to quantify, their determination provides a stumbling block for the analyst. However, since it is impossible to cover all scenarios that might exist and to calculate the probability of such occurrence, including the rare event in the general model will not increase our understanding of reality and will limit the clarity of the model.

Random Nature of the Monte Carlo Analysis.

Integral to any Monte-Carlo analysis is the generation of random numbers. Similar to rolling dice, the software has a 'random number generator' that produces a random sequence of numbers. Two main forms of sampling are Random Sampling (also called Monte Carlo Sampling) and Latin Hypercube sampling. Random or Monte Carlo sampling will evaluate the probability distributions in a purely random fashion, and is useful in trying to get the model to imitate a random sampling from a population or for doing statistical experiments. However, the randomness of this sampling suggests that, unless a very large number of iterations are performed, it will over-sample some parts and under-sample other parts of the distributions. Because for nearly all risk analysis modeling exploration of the distribution extremes (the "tails") is important, exact reproduction of the contributing distributions of the model becomes essential.

Latin Hypercube sampling (LHS) addresses this issue by providing a sampling method that appears random but that also guarantees to reproduce the input distribution with much greater efficiency that the random sampling. LHS uses a technique known as stratified sampling without replacement. It breaks the probability distribution into 'n' intervals of equal probability, where 'n' is the number of iterations to be performed on the model. Then, at random, one sample is drawn from each section, forcing, this way, an equal-chance representation of all the portions of the distribution. The Latin Hypercube method leads to a predictable uniformity of the sampling of the distribution.

For More Information

Use of Probabilistic Techniques (Including Monte Carlo Analysis) in Risk Assessment, Memorandum from the Office of the Administrator, U.S. Environmental Protection Agency, May 15, 1997

Policy for Use of Probabilistic Analysis in Risk Assessment at the U.S. Environmental Protection Agency. U.S. EPA, Office of Research and Development, May 15, 1997. (http://www.epa.gov/ncea/mcpolicy.htm)

Vose, David. Quantitative Risk *Quantitative Risk Assessment: a Guide to Monte-Carlo Simulation Modeling*. John Wiley and Sons (1996)